Special Report 89-20

June 1989

US Army Corps of Engineers
Cold Regions Research & Engineering Laboratory

Comparisons of low concentration measurement capability estimates in trace analysis

Method Detection Limit and Certified Reporting Limit

Clarence L. Grant, Alan D. Hewitt and Thomas F. Jenkins

Prepared for
U.S. ARMY TOXIC AND HAZARDOUS MATERIALS AGENCY
REPORT CETHA-TE-SR-88340

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188
Exp. Date: Jun 30, 1986

1a. REPORT SECURITY CLASSIFICATION Unclassified	1B. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Special Report 89-20	5. MONITORING ORGANIZATION REPORT NUMBER(S) CETHA-TE-SR-88340
6a. NAME OF PERFORMING ORGANIZATION U.S. Army Cold Regions Research and Engineering Laboratory 6b. OFFICE SYMBOL (if applicable) CECRL	7a. NAME OF MONITORING ORGANIZATION U.S. Army Toxic and Hazardous Materials Agency
6c. ADDRESS (City, State, and ZIP Code) 72 Lyme Road Hanover, N.H. 03755-1290	7b. ADDRESS (City, State, and ZIP Code) Aberdeen Proving Ground, Maryland 21010-5401
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (if applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Comparisons of Low Concentration Measurement Capability Estimates in Trace Analysis: Method Detection Limit and Certified Reporting Limit

12. PERSONAL AUTHOR(S)
Grant, Clarence L., Hewitt, Alan D. and Jenkins, Thomas F.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
	FROM _____ TO _____	June 1989	22

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Atomic absorption spectroscopy; Measurement capability measurements; Data analysis; Trace analysis; Liquid chromatography

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Two large data sets were obtained over a four-day period for graphite furnace atomic absorption spectroscopic measurement of copper (Cu) and reversed-phase high performance liquid chromatographic determination of dinitrobenzene (DNB) at a number of concentrations near the lower limit of measurement. Low concentration measurement capability estimates for each analyte were obtained using the Environmental Protection Agency's MDL (method detection limit) protocol and the U.S. Army Toxic and Hazardous Materials Agency's CRL (certified reporting limit) protocol. For DNB, analytical variance was found to be homogeneous over the concentration range examined and MDL estimates were independent of concentration over the range of concentration examined. MDL estimates varied by as much as a factor of three from day to day, emphasizing the uncertainty in these estimates. CRL estimates varied to about the same extent and were numerically quite similar to MDLs when equivalent α and β risks were used. For Cu, analytical variance was found to be proportional to concentration. Thus CRL estimates were very dependent on the concentration range examined. MDLs were less sensitive to this problem. Recommendations regarding the choice of target reporting limits for the CRL protocol were made. The influence of risk assumptions on both MDL and CRL estimates was examined and recommendations for modifications to both procedures made to incorporate an operational β-risk appropriate to the problem at hand. A case was made for using outlier tests to edit data used to estimate low concentration measurement capabilities.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Thomas Jenkins	603-646-4100	CECRL-RC

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

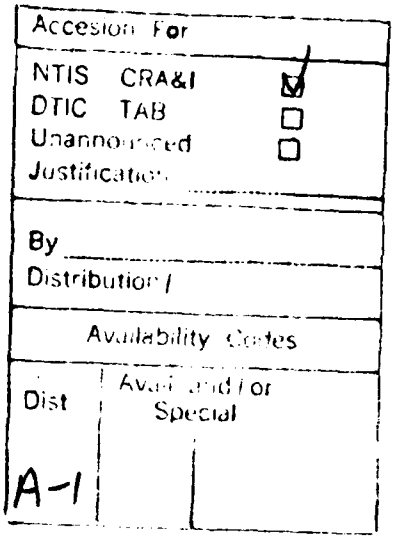
SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

PREFACE

This report was prepared by Dr. Clarence L. Grant, Consultant, Alan D. Hewitt, Research Physical Scientist, and Thomas F. Jenkins, Research Chemist, of the Geochemical Sciences Branch, Research Division, U.S. Army Cold Regions Research and Engineering Laboratory. Funding for this research was provided by the U.S. Army Toxic and Hazardous Materials Agency, Aberdeen Proving Ground, Maryland (R-90 Multi-Analytical Services), Martin H. Stutz, Project Monitor.

The authors gratefully acknowledge Dr. C.F. Bauer, University of New Hampshire and M.E. Walsh, CRREL, for their technical review of this report. The authors also acknowledge the technical assistance provided by P.W. Schumacher and P.H. Miyares during the experimental portion of the project.



CONTENTS

	Page
Abstract	i
Preface	ii
Introduction	1
Theory	2
Method Detection Limit	2
Certified Reporting Limit	3
Experimental	4
RP-HPLC determination of 1,3-dinitrobenzene	4
GFAA determination of copper	6
Results and discussion	7
MDL estimates for the DNB data set	7
CRL estimates for DNB data set	10
MDL estimates for Cu data set	12
CRL estimates for Cu data set	13
Conclusions and recommendations	14
Literature cited	15
Appendix A: Manual peak height measurements of chromatograms used to estimate DNB concentrations	17
Appendix B: Integrator peak height measurements of chromatograms used to estimate DNB concentrations	19
Appendix C: Graphite furnace atomic absorption measurements of copper concentrations	21

ILLUSTRATIONS

Figure

1. Graphical illustration of Method Detection Limit (MDL)	2
2. Relationship of CRL to criterion of detection	3
3. Graphical illustration of the relationship of the Certified Reporting Limit	4
4. Chromatograms of sample and standard showing separation of DNB from other nitroaromatics and nitramines	5

TABLES

Table

1. Results of linearity testing for DNB	5
2. Preparation of DNB test solutions for each of four days	6
3. Preparation of copper test solutions for each of four days	7
4. Mean found concentrations and standard deviations for manual peak height estimates of DNB	8
5. Mean found concentrations and standard deviations for integrator peak height estimates of DNB	9
6. Method detection limit estimates for manual peak height measurements of DNB concentrations	9

Figure Page
7. Variation of MDL estimates for DNB with changes in α- and β-risks 10
8. Certified reporting limit estimats for manual peak height measurements of DNB concentrations 11
9. Mean found concentrations and standard deviations for GFAA estimates of Cu 12
10. Method detection limit estimates for graphite furnace atomic absorption measurements of copper concentrations 13
11. Certified reporiting limit estimates for graphite furnace atomic absorption copper determinations 13

Comparisons of Low Concentration Measurement Capability Estimates in Trace Analysis

Method Detection Limit and Certified Reporting Limit

CLARENCE L. GRANT, ALAN D. HEWITT AND THOMAS F. JENKINS

INTRODUCTION

Reliable estimation of very low analyte concentrations in various sample types has commanded the attention of analytical chemists for many years. Terms such as detection limit, method detection limit, lower limit of reliable assay measurement, limit of quantitation, certified reporting limit, and many others have been introduced to describe this characteristic of analytical procedures. A recently published book (Currie 1988) contains an extensive literature review, historical information and descriptions of fundamentals and applications.

Achieving low concentration measurement capability is clearly required in many situations. However, specific requirements can vary widely and, therefore, a single experimental strategy is unlikely to meet all needs. A manufacturer of ultrapure material may require impurity reduction below a specified concentration for the product to be marketable. Here the quality control protocol will emphasize a narrow range of concentrations around the specification because any concentration above this value necessitates repurification. In contrast, if the purification procedure is being experimentally optimized, calibration over a somewhat wider concentration range may be required to permit accurate assessment of the effect of variable manipulation.

An environmental survey of a potential toxic waste site usually dictates the need to quantitate over a wide concentration range. To plan an effective remedial program, high concentrations must be determined with reasonable accuracy. Simultaneously, concentrations at the regulatory or action level must be measured in order to define the geographic boundaries of contamination. Because calibration for many analyses is costly in both time and money, most laboratories attempt to satisfy both requirements with a single protocol. Not surprisingly, this requires some degree of compromise. The alternative would be to bracket analyte responses in samples with calibration standards. Because more time would be devoted to the analysis of calibration standards, the number of samples analyzed would need to be reduced to keep costs manageable. Since sampling usually contributes the largest amount of uncertainty in environmental surveys, reducing sample numbers is an unattractive option.

The major point of the above discussion is to emphasize that detection capability is not a fundamental property of procedures nor is it a constant. Instead it can be "managed" to a significant extent based on the experimental protocol (number of standards and blanks analyzed, distribution of standards, time span covered by standards measurements, number of replications of standards and unknowns, etc.). The risks chosen for statistical decisions also play a major role. While most detection limit definitions have only provided protection against false positives (type I or α risks), there is increased recent emphasis on the need to protect against false negatives (type II or β risks) (Currie 1968 and 1988, Hubaux and Vos 1970, Kirchmer 1983, Wernimont 1985, Clayton et al. 1987). Also receiving more attention of late is the difference between detection (qualitative) and quantitation (Crummet et al. 1980). Clearly, the levels of assurance chosen for specifying detection criteria will have a substantial impact on their magnitude.

Analytical precision is another major contributor to the size of these estimates; better precision means better low concentration measurement capability, other things being equal. Here again there is a complex relationship between precision estimates and the specific methodology, the experience and attitude of the analyst, the condition of the apparatus, and the laboratory cleanliness. In view of all this, it should not be surprising that reported estimates of detection and quantitation limits vary widely both between and within laboratories.

Two common estimates of low concentration measurement capability for environmental studies

are 1) the Method Detection Limit (MDL) specified by the U.S. Environmental Protection Agency (Glaser et al. 1981, Federal Register 1984), and 2) the Certified Reporting Limit (CRL) specified by the U.S. Army Toxic and Hazardous Materials Agency (USATHAMA 1987). At a recent conference (Maskarinec and Holladay 1987) and elsewhere, it has been suggested that these two procedures yield widely divergent estimates for the same analyses.

The purpose of this study is to examine this question by extracting such estimates from two extensive data sets. The first was the determination of 1,3-dinitrobenzene (DNB) by reversed-phase high performance liquid chromatography (RP-HPLC) (Jenkins et al. 1988) and the second was the determination of copper by graphite furnace atomic absorption (GFAA). These methods were chosen because prior data showed that the precision of the DNB measurements was relatively constant in the concentration range used, whereas precision for Cu showed a regular dependence on concentration. Besides comparing MDL with CRL under identical experimental conditions, the program was designed to demonstrate the effects of variations in α and β risks and changes in experimental protocols.

THEORY

Method Detection Limit (MDL)

The MDL is defined as the "minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero and is determined from analysis of a sample in a given matrix containing the analyte" (Federal Register 1984). After estimating the MDL from instrumental responses and prior experience of the analyst, either reagent water or another sample matrix is spiked (if necessary) to give an analyte concentration that is one to five times the estimated MDL. A minimum of seven replicate aliquots are processed through the entire analytical procedure. If a blank is required, a separate blank measurement is obtained for each sample and the average blank measurement is subtracted from each sample measurement.

Computation of the estimated MDL is accomplished by multiplying the standard deviation of the replicate measurements by the appropriate one-sided t-value corresponding to $n-1$ degrees of freedom and a 99% confidence level. It is assumed that variances are reasonably homogeneous and that error distributions approximate normal in the region from the blank to five times the MDL. Although both of these assumptions are frequently violated to a small extent, the error produced is usually acceptable when compared to overall uncertainties. No allowance is made in the MDL estimate for any error contribution by the blank. It is also important to note that while the risk of false positives is only 1% ($\alpha = 0.01$), the risk of false negatives is 50% ($\beta = 0.50$) for a sample with a true concentration equal to the current MDL (Fig. 1). Kirchmer (1983, 1988) has chosen to call this quantity the criterion of detection.

To reduce the risk of false negatives requires incorporation of a realistic β risk factor in setting detection or reporting limits (Fig. 1). The size of α and β risks can be varied to fit the requirements of the problem at hand (Wernimont 1985, Dixon and Massey 1969). For example, the possible revised MDL shown in Figure 1 has a β risk equal to the

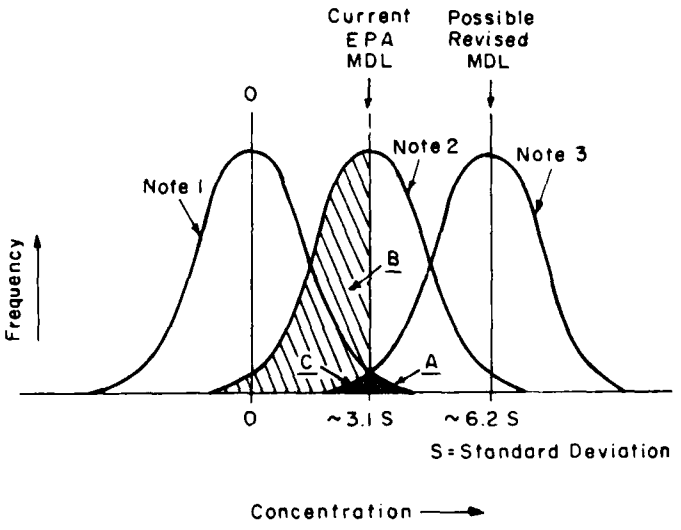


Figure 1. Graphical illustration of Method Detection Limit (MDL).
Note 1. Distribution of blank measurements.
Note 2. Distribution of measurements with a mean concentration equal to the current MDL with α = 0.01 (shown as area A), i.e., risk of claiming detection when true concentration is zero, and β = 0.50 (shown as area B), i.e., risk of claiming absence when true concentration equals MDL.
Note 3. Distribution of measurements with a mean concentration with β = 0.01 (shown as area C). The α risk for this possible revised MDL is approximately zero. Risks other than 0.01 could be used for both α and β.

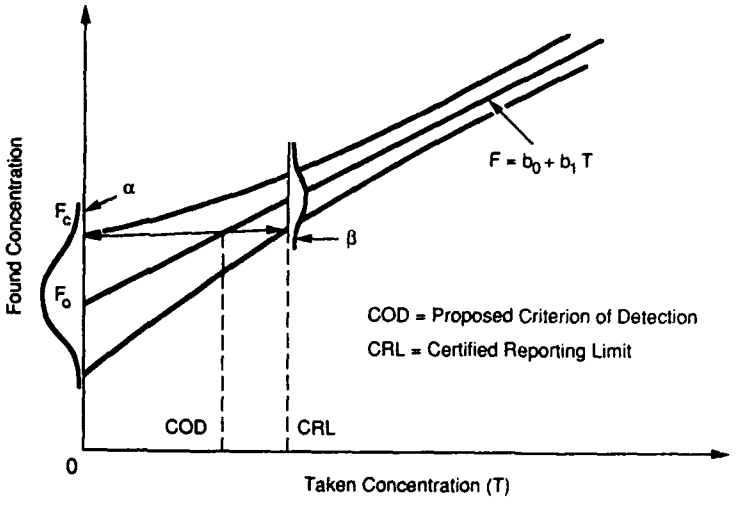


Figure 2. Relationship of CRL to criterion of detection (COD).
Note 1. CRL is the value of T corresponding to a point on the lower confidence band where the value of F equals the value of F at T= 0 on the upper confidence band.
Note 2. The curved confidence bands represent the joint uncertainties in the slope (b_1) and intercept (b_0). Current USA THAMA requirements specify α = β = 0.05, but other values could be used.

initially chosen α risk (1%) but there is nothing to prevent choosing other values for these risks. We believe that an experimental value below the criterion of detection (current MDL) should be reported as "not detected, less than the revised MDL" in order to obtain protection against false negatives. Values between the criterion of detection (current MDL) and the revised MDL could be reported in parentheses with a notation to explain that these estimates are less reliable than those above the revised MDL. Unfortunately this system is slightly cumbersome in terms of data processing, which explains why most detection limit estimates are used as a cutoff for numerical values with everything below this value reported as not detected. When used in this manner, the effective β risk is always 50%.

In the *Results and Discussion* section of this report we will demonstrate the effects on MDL estimates as these risks are varied. The effect of variance inhomogeneity will also be examined.

Certified Reporting Limit (CRL)

The CRL specified by USATHAMA (1987) is extracted using confidence bands as described by Hubaux and Vos (1970). A target reporting limit (TRL) is chosen based on method capability and data requirements and spike additions of the analyte are made at concentrations ranging from 0.5 TRL to 10 TRL. A linear least-squares regression model of the form $F = b_0 + b_1 T$ is obtained from a plot of *found* concentrations (*F*) vs *taken* concentrations (*T*). The data for the plot are obtained on four separate days and then are pooled. Thus, day-to-day calibration error is included in the pooled standard deviation estimate. This is one distinct difference between MDL and CRL estimates.

The least-squares linear model is then tested against the theoretically expected linear model through the origin. When lack-of-fit tests indicate departure from a linear model, the highest concentration values may be sequentially truncated until linearity is achieved, except that a minimum of three concentration values must remain. A CRL estimate is extracted using two-tail confidence bands with (α = β = 0.05) as shown in Figure 2. The standard USATHAMA protocol requires that at least one of the tested concentrations must be below the CRL; otherwise the lowest tested concentration is the CRL. In practice, however, when the CRL is used as a cutoff point and all values below the CRL are reported as "not detected," the operational β risk becomes 50% just as it is for an MDL used in this fashion. One way to approach this problem would be through the inclusion of a "criterion of detection" (COD) concentration similar to that suggested for the MDL (Fig. 2). Thus, the "criterion of detection" would be estimated from the intersection of the horizontal line corresponding to *F* at *T* = 0 on the upper confidence band with the best fitting model (Fig. 2). Values below this concentration of *T* would be reported as "not detected, less than CRL." Values between the COD and CRL would be reported in parentheses as described for MDL.

Even when the original data set is adequately described by a linear model, the CRL may be higher than required for the intended use of the data. This situation could arise as a consequence of a large pooled standard deviation where the major uncertainty is produced at the higher concentrations. Hubaux and Vos (1970) note that homogeneity of variance is assumed and that assumption may not hold for some analyses. The most common

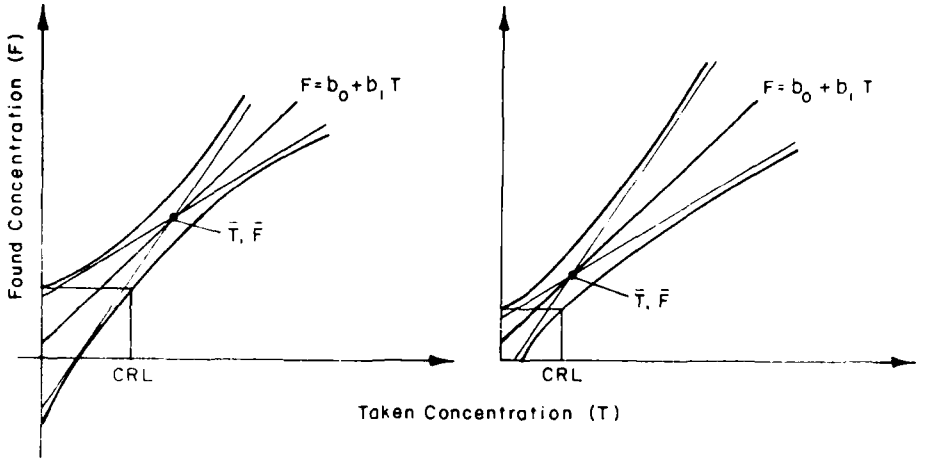


Figure 3. Graphical illustration of the relationship of the Certified Reporting Limit (CRL) to the centroid $\overline{T}, \overline{F}$.

departure is when the variance increases with concentration. This situation can be dealt with in several ways. One way is to confine all measurements to a very low concentration range but this defeats the objective of reliable quantitation over a wide concentration range. This problem could also be addressed by adding at least one more standard to the calibration at 0.25 TRL or perhaps an even lower concentration. Another possibility is to perform many replicate measurements at low concentration and very few at high concentration but this also fails to satisfy the requirement of accurate calibration over an extended range. Oppenheimer et al. (1983), recommend weighting, which seems like an attractive option but has not yet been widely employed. In the USATHAMA protocol, this problem is dealt with by truncation of the data set but this procedure suffers the disadvantage of being arbitrary.

Hubaux and Vos (1970) also suggest that, ideally, the mean concentration of all standards should be as close to the detection limit as possible. Since the point $\overline{T}, \overline{F}$ is the centroid of rotation for uncertainty in the slope, a low value for this point reduces the extent of extrapolation required and thereby reduces the CRL (Fig. 3). The options for reducing the size of $\overline{T}, \overline{F}$ are similar to those for reducing the pooled standard deviation. When truncating according to USATHAMA requirements, the slope of the least-squares linear regression line after each truncation must not differ by more than 10% from the slope for the total data set if the original linear model was an adequate fit to the data.

In the *Results and Discussion* section of this report we will examine the effects of truncation on CRL estimates for two data sets where one exhibits relatively constant variance and the other shows dependence of variance on concentration. The effect of choosing different TRLs will also be considered along with variations in α and β risks (Wernimont 1985). Differences and similarities in CRL and MDL estimates obtained from the same data sets will be related to variations in underlying assumptions.

EXPERIMENTAL

RP-HPLC Determination of 1,3-dinitrobenzene

Instrumental method

All RP-HPLC determinations for 1,3-dinitrobenzene (DNB) were obtained on a modular instrument consisting of a Perkin-Elmer Series 3 pump, a Dynatech Model LC-241 Autosampler equipped with a 100-µL sample loop, a Perkin-Elmer variable wavelength detector operated at 254 nm, a Hewlett-Packard 3393 digital integrator and a Linear Model 500 strip chart recorder. The method involves dilution of the aqueous samples 1/1 (V/V) with methanol, filtration through a 0.5-µm Millex-SR filter and determination on a 25-cm × 4.6-mm (5-µm) LC-18 column (Supelco), eluted with 1.5 mL/min of 1/1 (V/V) methanol–water (Jenkins et al. 1988). Chromatograms showing separation of DNB from other nitroaromatics and nitramines are shown in Figure 4.

Chemicals

Analytical standards for DNB were prepared from Standard Analytical Reference Material (SARM) obtained from the U.S. Army Toxic and

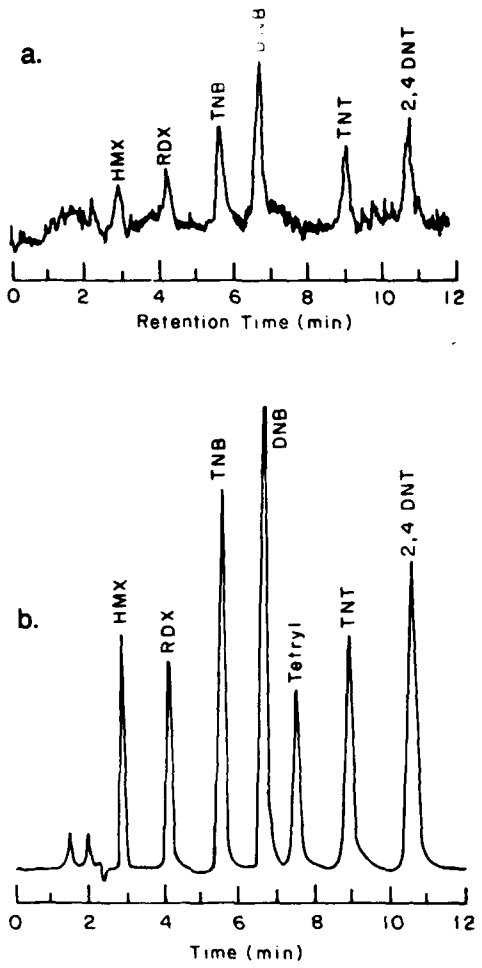


Figure 4. Chromatograms of sample (a) and standard (b) showing separation of DNB from other nitroaromatics and nitramines.

Hazardous Materials Agency. SARM was dried to constant weight in a vacuum desiccator in the dark.

Baker HPLC grade methanol was used to dilute samples, prepare standards and serve as the RP-HPLC eluent. Mallinckrodt ChromAR grade acetonitrile was used in preparation of the stock standard. Water was prepared on a Milli-Q type 1 Reagent Grade Water System (Millipore Corporation). Water and methanol were combined in equal proportions and vacuum filtered through a Whatman CF-F microfiber filter to remove particulates and degas the eluent.

Calibration

The stock calibration standard for DNB was prepared by accurately weighing out approximately 100 mg of dried SARM and diluting to volume with acetonitrile in a 250-mL volumetric flask. A combined analyte stock standard was prepared by combining 2.00 mL of the DNB stock standard with similar volumes of stock standards of HMX, RDX, TNB, tetryl, TNT and 2,4-DNT in a 200-mL volumetric flask and diluting to volume with methanol. A diluted combined stock standard was prepared by placing 50.0 mL of the combined stock standard in a 200-mL volumetric flask and diluting to volume with methanol. A series of eight calibration standards were prepared from the diluted standard with DNB concentrations ranging from 2.57 to 514 µg/L. Prior to analysis, all standards were diluted 1/1 (V/V) with water.

Table 1. Results of linearity testing for DNB.

Concentration (µg/L)	Detector response (Integrator units) Replicate A	Replicate B
2.57	568	0
6.14	621	673
10.3	1303	1231
25.7	1962	2140
51.4	3680	3556
103	7712	7406
257	18317	19067
514	30334	35538

Regression equation response = 66.195 (concentration).

Peak heights resulting from duplicate injections of these eight standards in random order were obtained from the digital integrator (Table 1). Lack-of-fit testing of the data indicated that a linear model with zero intercept adequately described the relationship between detector response and analyte concentration at the 95% confidence level. Similar results were obtained when peak heights were measured manually from the strip chart output. These results were consistent with earlier tests described elsewhere (Jenkins et al. 1988, Jenkins and Walsh 1987). Thus for daily calibration, replicate injections of a single standard with a DNB concentration of 256 µg/L were used for detector calibration.

Preparation of test samples

On each of four days a 25.0-mL aliquot of the combined stock standard was diluted to 250 mL with methanol in a volumetric flask. This solution is referred to as the 10 TRL sample. Further dilutions of this sample were made as shown in Table 2. Although the USATHAMA protocol only requires solutions corresponding to 0.5, 1, 2, 5, and 10 times the TRL, several extra dilutions were included to permit calculations of CRLs based on different choices for the TRL.

Table 2. Preparation of DNB test solutions for each of four days.

Solution	Volume of 10 TRL (mL)	Total volume (mL)	DNB concentration (µg/L)	Replicates analyzed per day
10 TRL			411	2
5 TRL	50	100	205	2
2.5 TRL	25	100	103	2
2 TRL	50	250	82.2	10
1 TRL	25	250	41.1	10
0.5 TRL	10	200	20.5	10
0.25 TRL	25 mL of 1 TRL	100	10.3	10
0.125TRL	25 mL of 1 TRL	200	5.1	10
Blank			0	10

On each of four days the numbers of replicates shown in Table 2 were analyzed in random order. For the 0.125 TRL to 2 TRL levels, 10 replicates were analyzed to enable calculation of an MDL at levels of 5.1, 10.3, 20.5, 41.1 and 82.2 µg/L on each of the four days. In order to do CRL calculations using the USATHAMA protocol, a target value (TRL) is chosen and at least one replicate of 0.5, 1, 2, 5 and 10 times the TRL must be analyzed on each of the four days. We chose to use duplicates from each day at each level, and thus two replicates were analyzed at the 2.5, 5 and 10 TRL level each day. Where more values were available, we randomly selected values from the group of measurements obtained each day. In all cases, peak heights were obtained both manually from the strip chart output and automatically using the digital integrator. This matrix of results allowed us to obtain CRL estimates for TRLs of 10.3,* 20.5 and 41.1 µg/L.

GFAA determination of copper

Instrumental Method

Aqueous Cu determinations were performed on a Perkin-Elmer model 403 atomic absorption spectrometer coupled with a HGA-2200 graphite furnace controller. Instrumental response was monitored by peak height measurements of tracings obtained with a Linear model 500 strip chart recorder. All manual injections were made with a 10-µL fixed volume Eppendorf syringe. Analyses were performed employing pyro-coated graphite tubes with the following furnace program: dry for 15 sec at 110°C, char for 10 sec at 850°C, and atomize for 3 sec at 2700°C with an argon gas flow of 30 cm^3/min. Chart recorder gain was 0.050 or 0.100 AFS (absorbance for full-scale deflection) with no signal damping.

Chemicals and Materials:

Copper solutions for standards and samples were made by diluting a 1000-mg/L certified atomic absorption reference solution (Fisher Scientific Corp.). The sample and standard matrix consisted of reagent grade water (Milli-Q from Millipore) acidified to 0.2% V/V with G. Frederick Smith (GFS) distilled HNO_3. All diluted solutions were stored in low density polyethylene (Nalgene) bottles that had been water rinsed, soaked for 48 hours in 10% V/V reagent grade HNO_3, rinsed with reagent grade water, filled with 1% V/V GFS HNO_3, emptied and rinsed with reagent grade water and dried prior to use. Individual sample bottles were reused with solutions of the same Cu concentration after being rinsed with reagent grade water and dried between daily runs. Pipette tips were soaked in concentrated HNO_3 for several days and then rinsed with reagent grade water and dried. Sample preparation and material cleaning were performed in a clean air station inside a class 100 clean room.

Calibration

A 10.0-mg/L stock standard was prepared by making a 100-fold dilution of the Fisher Scientific 1000 mg/L reference standard. Standards of 12.0, 6.00, 3.00, and 1.20 µg/L Cu were prepared daily (along with the samples) to establish sensitivity. The instrumental settings employed produced a marked response to the furnace program without aqueous sample introduction. This "furnace blank" was constant throughout the experiment and was assumed to result from light emitted from the

* When the TRL was 10.3 µg/L, measurements of 0.5, 1, 2, 4 and 10 TRL were obtained.

graphite tube at 2700°C. After subtraction of the furnace blank, lack-of-fit testing of standard calibration curves showed that a linear model with zero intercept adequately described the calibration data at the 95% confidence level.

Target level

The target level (1.2 µg/L) was estimated based on 15 times the measured deflection of the baseline noise present, and the measured sensitivity. The sensitivity obtained was in good agreement with the manufacturer's suggested instrumental capabilities.

Preparation and analysis of samples

Analysis of the 48 individually prepared aqueous Cu samples (Table 3) was performed in random order on four separate days. Fresh standards and samples were prepared daily by first diluting the 10.0-mg/L stock standard to 200 µg/L. The 200-µg/L standard was used to prepare the 12.0- and 2.40- µg/L solutions. Further dilutions were prepared from the 12.0- and 2.40- µg/L solutions as shown in Table 3. All samples and standards were prepared and analyzed within a five-hour period. Two injections of each calibration standard (12.0, 6.00, 3.00, and 1.20 µg/L) were made in random order before and after running the 48 samples. All 16 points were employed to establish the calibration curve. The accuracy of the calibration curve was verified with a certified EPA trace metal standard. The samples and standards were analyzed by a single injection, in random order. During the analysis of samples, seven furnace blanks were also obtained and the average blank was subtracted from all samples. These determinations could not be performed blindly since the recorder gain had to be changed for concentrations greater than 3.00 µg/L.

MDL estimates were obtained for Cu concentrations of *0.15, 0.30, 0.60, 1.20* and *2.40* µg/L on each of four days (total of 20 estimates). CRL estimates were obtained for TRL values of 0.30, 0.60 and 1.20 µg/L using two values from each day as described under the DNB analysis section.

It is important to recognize that this experimental design was developed specifically to permit direct comparison of MDL and CRL estinates. It varies from routine practice for MDL estimation in that only one or two sets of seven (minimum) replicate measurements would normally be made. Similarly, the design for CRL measurements would not normally entail as many different concentrations as used here beacause only one TRL would be chosen. In other aspects, however, the CRL design does simulate normal practice. Clearly, the CRL procedure is intended to produce several bits of information, whereas the MDL procedure is designed specifically for this one estimate. In the CRL procedure it is possible to 1) compare within-day and between-day variability (CRL includes both), 2) determine the linear range of calibration, 3) estimate precision and accuracy over the full linear range, and 4) obtain an estimate of low concentration measurement capability (CRL). For this study, only the CRL estimate is used.

Table 3. Preparation of copper test solutions for each of four days.

Solution	*Vol. of 10 TRL (mL)*	*Final vol. (mL)*	*Copper conc. (µg/L)*	*Replicates analyzed per day*
10 TRL		1.00	12.0	2*
5 TRL	1.00	2.00	6.00	2*
2.5 TRL	1.00	4.00	3.00	2*
2 TRL	Volume of 2 TRL	1.00	2.40	7
1 TRL	1.00	2.00	1.20	7*
0.5 TRL	1.00	4.00	0.600	7
0.25 TRL	0.500	4.00	0.300	7
0.125TRL	0.250	4.00	0.150	7
Blank	0.000	1.000	0.000	7

* An additional aliquot was prepared for the standards.

RESULTS AND DISCUSSION

MDL estimates for the DNB data set

Complete data sets for both manual and integrator peak height estimates of DNB concentrations are given in Appendices A and B. Before computing variances, the data were examined for the presence of individual aberrant values using Dixon's (1953) test at the 5% significance level, and for multiple outliers according to the test described by Grubbs (1969), again at the 5% significance level.

Six single outliers were identified in the manual peak height estimates. For the integrator data a total of five outliers were found with two of those residing within one set of 10 replicates. This was the only case of multiple outliers in a single set of measurements. Surprisingly, there was no correlation of location of the outliers in the manual set compared to the integrator set even through the estimates were extracted from the same chromatograms.

Variance homogeneity testing was also conducted at the 5% significance level using Cochran's

range comparison (Youden and Steiner 1975). Significant heterogeneity was present when the entire data sets were tested but exclusion of the 11 outliers eliminated this heterogeneity. Since the 11 outliers account for only 2.9% of the results for concentrations of 5.1, 10.3, 20.5, 41.1, and 82.2 µg/L, comparisons were based on the edited data (although computations were also made for the unedited data). No editing was performed on the duplicate measurements for the three highest concentrations.

The elimination of outliers by a statistical procedure is clearly controversial, and unfortunately there is no completely unequivocal solution to this problem. However, when a single result in a set of otherwise typical random measurements is so aberrant that the variance is grossly elevated, we believe that value should be deleted. After all, the objective of data collection is to represent fairly the capability of a procedure. For DNB the standard deviations after editing for the 10 sets of replicates containing an outlier (one set had two outliers) were less than half the original estimates. Furthermore, the standard deviations (s) for the 10 edited data sets were now in excellent agreement with the 30 standard deviation estimates from unedited data sets. Consequently, we believe that this very small amount of editing was not only proper, but that it should represent routine practice in such studies.

Mean found concentrations ($\bar{X}$) and standard deviations of individual measurements are summarized in Table 4 for manual peak height estimates and in Table 5 for integrator estimates. In addition to the separate daily estimates, means and standard deviations were computed for the total data sets at each concentration. Finally, pooled standard deviations were calculated by combining daily variances; i.e., systematic day-to-day variations were excluded. For the three highest concentrations where only two measurements were recorded each day, $\bar{X}$, and s estimates were based on the total of eight values at each concentration.

In general, mean found concentration estimates extracted from the manual peak height data were in excellent agreement with taken concentrations. However, there was some evidence of a very small positive bias at low concentrations and a small negative bias at high concentrations. Daily estimates of s for the five lowest concentrations ranged from 0.94 to 4.31 µg/L. Estimates based on the total data ranged from 1.95 to 3.22 µg/L for the five lowest concentrations. For the concentrations above this range, only the highest one (410.8 µg/L) showed a significant increase in s. Note also that the total and pooled standard deviations were nearly identical, indicating that day-to-day variation in these data was similar in magnitude to within-day variation. This conclusion was further confirmed by analysis of variance, which showed no significant differences ($\alpha = 0.05$) in daily mean responses.

In comparison, the integrator peak height esti-

Table 4. Mean found concentrations and standard deviations for manual peak height estimates of DNB (edited data).

Data set	Concentration (µg/L) Taken	Found ($\bar{X}$)	Degrees of freedom (d.f.)	Standard deviation (s)
Day 1	5.1	5.6	9	3.01
Day 2	5.1	6.5	9	1.82
Day 3	5.1	6.5	9	3.45
Day 4	5.1	5.7	9	4.31
Total*	5.1	6.3	39	3.22
Pooled**	—	—	36	3.22
Day 1	10.3	10.6	9	1.52
Day 2	10.3	12.3	9	1.71
Day 3	10.3	11.5	9	2.22
Day 4	10.3	11.1	9	2.12
Total*	10.3	11.4	39	1.95
Pooled†	—	—	36	1.92
Day 1	20.5	21.8	9	1.42
Day 2	20.5	21.8	9	2.10
Day 3	20.5	22.1	8	0.94
Day 4	20.5	20.8	9	3.72
Total*	20.5	21.6	38	2.29
Pooled†	—	—	35	2.33
Day 1	41.1	41.3	9	1.65
Day 2	41.1	43.1	9	3.10
Day 3	41.1	41.6	8	2.69
Day 4	41.1	40.7	8	2.83
Total*	41.1	41.7	37	2.67
Pooled†	—	—	34	2.62
Day 1	82.2	81.3	8	1.84
Day 2	82.2	80.0	6	2.84
Day 3	82.2	82.0	5	2.09
Day 4	82.2	81.0	5	1.90
Total*	82.2	81.1	27	2.22
Pooled†	—	—	24	2.19
Total**	102.7	101.0	7	4.27
Total**	205.4	201.0	7	3.60
Total**	410.8	398.0	7	6.83

* The total standard deviation is based on combining the data from all four days into a single set.

† The pooled standard deviation combines the four variances from each day but with day-to-day variations excluded.

** For the three highest concentrations where duplicates were run each day, only the total standard deviation is reported.

Table 5. Mean found concentrations and standard deviations for integrator peak height estimates of DNB (edited data).

Data set	Concentration (µg/L) Taken	Found ($\bar{X}$)	Degrees of freedom (d.f.)	Standard deviation (S)
Day 1	5.1	4.8	9	5.33
Day 2	5.1	11.2	9	6.93
Day 3	5.1	7.5	9	4.49
Day 4	5.1	8.5	8	5.47
Total*	5.1	8.0	38	5.89
Pooled**	—	—	35	5.63
Day 1	10.3	12.5	9	7.01
Day 2	10.3	15.7	9	5.01
Day 3	10.3	18.1	9	5.37
Day 4	10.3	11.8	9	7.23
Total*	10.3	14.5	39	6.52
Pooled†	—	—	36	6.23
Day 1	20.5	28.7	8	3.52
Day 2	20.5	24.7	7	3.14
Day 3	20.5	24.8	9	4.91
Day 4	20.5	27.4	9	6.31
Total*	20.5	26.4	36	4.68
Pooled**	—	—	33	4.73
Day 1	41.1	44.5	8	1.90
Day 2	41.1	49.2	9	4.66
Day 3	41.1	47.5	8	8.58
Day 4	41.1	48.5	8	8.85
Total*	41.1	47.5	38	6.85
Pooled†	—	—	35	6.74
Day 1	82.2	89.8	9	6.87
Day 2	82.2	89.0	6	9.16
Day 3	82.2	81.5	6	4.31
Day 4	82.2	87.2	6	4.32
Total*	82.2	87.2	30	6.99
Pooled†	—	—	27	6.53
Total**	102.7	109.0	7	7.89
Total**	205.4	203.0	7	10.7
Total**	410.8	399.0	7	11.3

* The total standard deviation is based on combining the data from all four days into a single set.
† The pooled standard deviation combines the four variances from each day but with day-to-day variations excluded.
** For the three highest concentrations where duplicates were run each day, only the total standard deviation is reported.

mates show similar trends but 1) the found concentrations are biased by a greater amount, especially those on the high side, and 2) the standard deviation estimates are larger by a factor of more than two than for the manual peak height data. Since both MDL and CRL are proportional to s, the integrator derived estimates will be more than twice as large as manual peak height estimates. Consequently, only the manual results will be used for MDL and CRL comparisons. We strongly recommend visual inspection of chromatograms when integrator data collection are employed. Otherwise, some spurious data that misrepresent the chromatograms will likely be collected, especially at very low response levels.

Table 6. Method detection limit (MDL)* estimates for manual peak height measurements of DNB concentrations. Estimates are based on t-values with $\alpha = 1\%$ and $\beta = 50\%$.

Concentration taken (µg/L)	MDL estimates (µg/L) Day 1	Day 2	Day 3	Day 4	Pooled†
5.1	8.5	5.1	9.7	12.1	8.0
10.3	4.3	4.8	6.3	6.0	4.7
20.5	4.0	5.9	2.7	10.5	5.7
41.1	4.7	8.7	7.8	8.2	6.4
82.2	5.2	8.9	7.0	6.4	5.5

* MDL = $t_{0.99}(S)$
† Pooled across the four days at each concentration
$t_{0.99}$(5 d.f.) = 3.365, $t_{0.99}$(6 d.f.) = 3.143, $t_{0.99}$(7 d.f.) = 2.998, $t_{0.99}$(8 d.f.) = 2.896, $t_{0.99}$(9 d.f.) = 2.821, $t_{0.99}$(24 d.f.) = 2.492, $t_{0.99}$(34 d.f.) = 2.443, $t_{0.99}$(35 d.f.) = 2.440, $t_{0.99}$(36 d.f.) = 2.437

In accordance with EPA recommendations (Federal Register 1984) MDLs have been calculated using the standard deviations for each set of daily replicates. Appropriate one-sided t-values giving an α-risk of 1% and a β-risk of 50% (Fig. 1) were multiplied times the corresponding estimates of s (Table 6). Although the specifications require s to be derived using analyte concentrations within 1 to 5 times the estimated MDL, we have included one higher concentration estimate (82.2 µg/L) because there was no evidence that s had increased above those for the lower concentration values. The MDLs for manually derived peak height estimates ranged from 2.7 to 12.1 µg/L. In contrast, comparable estimates for integrator-derived data ranged from 5.5 to 28.8 µg/L. MDLs for pooled standard deviations from manual peak height estimates (Table 6) range only from 4.7 to 8.0 µg/L. These values are lower than the daily estimates due to smaller t-values corresponding to the large degrees of freedom in pooled estimates of s.

It is interesting to compare the observed scatter of MDL values with the scatter we would predict from confidence limits around the day 1 estimate (MDL = 8.5 µg/L) for the 5.1-µg/L concentration. The chi-squared distribution at the 95% confidence level yielded a range of 5.8 to 15.5 µg/L. Inspection

Table 7. Variation of MDL estimates for DNB with changes in α- and β-risks. Assume nine degrees of freedom and s = 2.42 µg/L throughout.

α-risk (%)	β-risk (%)	MDL (µg/L)
1	50	6.84
1	10	10.2
1	5	11.3
1	1	13.2
5	50	4.43
5	10	7.65
5	5	8.59
10	50	3.34
10	10	6.39

of Table 6 shows no daily estimates above 15.5 µg/L but 7 of the 20 values were below the lower boundary of 5.5 µg/L. However, if the mean MDL for all of the day 1 data is used, the confidence limits become 3.7 to 9.7 µg/L. Now, one of the values is below the lower boundary and two exceed the upper boundary. The point here is to reinforce the notion that detection capability estimates are highly uncertain and that this uncertainty is greater when the amount of data collected is small.

Let us now examine the influence of risk assumptions on the size of MDL estimates. For this comparison we will use the pooled standard deviation estimate for the five lowest concentrations from Table 4 (s = 2.42 µg/L) but we will assume only nine degrees of freedom to correspond with a single set of 10 measurements on one day at one concentration. A composite t value representing the sum of the t-values for the chosen α- and β-risks were computed from the equation in Dixon and Massey (1969, p. 273):

$$2t = \left[z_{(1-\alpha)} + z_{(1-\beta)}\right]\left[1 + \frac{1.21\,(z_{(1-\alpha)} - 1.06)}{\text{d.f.}}\right] \qquad (1)$$

where z is the standard normal variable and d.f. is the degrees of freedom. It is claimed that the equation yields values accurate to within 0.5% when d.f. ≥ 9. As expected, MDL estimates for DNB increased considerably when β-risks were reduced to sizes comparable to the α-risks (Table 7). If the recommendations given in the *Theory* section were adopted, the values for α = β = 5% would be 4.43 µg/L for the COD and 8.59 µg/L for the MDL. Values between these limits would be reported in parentheses and values below 4.43 µg/L would be reported as "not detected < 8.59 µg/L." The MDL according to the current EPA definition is 6.84 µg/L. We will also refer to this tabulation when we compare MDL to CRL estimates.

Before leaving this topic, we should note that no provision has been made for uncertainty associated with a blank correction. For these DNB measurements, no blank corrections were necessary (blank not significantly different from zero). However, when blank corrections are required, the MDL estimates would increase by $\sqrt{2}$ to account for the added uncertainty attached to a difference calculation.

CRL estimates for DNB data set

CRL estimates were obtained from edited manual and integrator peak height measurements of DNB according to the USATHAMA (1987) protocol. Blank measurements were not employed in fitting the regression models. Three target reporting limits (TRL) were compared (10.3, 20.5 and 41.1 µg/L) in order to observe the effect on CRL estimates. In addition to the CRLs derived from full data sets (0.5, 1, 2, 5 and 10 TRL), two successive truncations of high concentration values were applied at each TRL.

Since duplicate values were randomly selected from daily data sets up to a concentration of 82.2 µg/L, this process was repeated five times to obtain a measure of the variability to be expected in CRL estimates. We must caution, however, that this variation is biased on the low side in some cases because the highest concentrations used in the regression models were only measured in duplicate each day. Thus, when the TRL was 10.3 µg/L, all five CRL estimates for the full data set used the same two values at the 103-µg/L concentration level (10 TRL). The five CRL estimates for the two truncations, however, were derived from randomly generated data over the entire concentration range represented. When TRL was 20.5 or 41.1 µg/L, only the CRL estimates from the second truncation are based entirely on randomly generated data.

CRL estimates for manual peak height measurements of DNB are summarized in Table 8. The estimates for integrator data exhibit similar trends but they are numerically much larger. Consequently, only the estimates from manual measurements

Table 8. Certified reporting limit (CRL)* estimates for manual peak height measurements of DNB concentrations.

Target reporting limit (μg/L)	*Full data set*	*Highest conc. deleted*	*Two highest conc. deleted*
10.3	8.2	8.5	8.3
10.3	9.2	10.0	9.9
10.3	8.5	7.6	6.0
10.3	8.3	9.4	9.9
10.3	8.0	7.2	5.9
	$\bar{X}$ = 8.4	$\bar{X}$ = 8.5	$\bar{X}$ = 8.0
	R† = 1.2	R = 2.8	R = 4.0
20.5	11.0	10.9	9.8
20.5	12.0	12.2	11.9
20.5	10.4	10.1	7.0
20.5	14.1	14.9	16.2
20.5	10.2	9.8	7.5
	$\bar{X}$ = 11.5	$\bar{X}$ = 11.6	$\bar{X}$ = 10.5
	R = 3.7	R = 5.1	R = 9.2
41.1	14.2	10.5	19.4
41.1	12.9	8.3	6.6
41.1	13.8	10.0	9.6
41.1	17.2	15.3	15.4
41.1	17.5	15.9	17.9
	$\bar{X}$ = 15.1	$\bar{X}$ = 12.0	$\bar{X}$ = 12.0
	R = 4.6	R = 7.6	R = 11.3

* Based on $\alpha = \beta = 5\%$ using USATHAMA (1987) computational protocol (Fig. 2).

† Range between high and low CRL estimate.

are tabulated. These values, which are based on $\alpha = \beta = 5\%$, show some important systematic trends:

1. As TRL increased, CRL estimates increased. This trend was independent of whether full or truncated data sets were used. Specifically, CRL ranged from 5.9 to 10.0 μg/L for TRL = 10.3 μg/L, from 7.0 to 16. 2 μg/L for TRL = 20.5 μg/L, and from 6.6 to 17.9 μg/L for TRL = 41.1 μg/L. The corresponding means were 8.3, 11.2, and 13.0 μg/L, respectively (Table 8). This behavior is thought to be caused primarily by the longer extrapolation required for slope confidence bands as the centroid of rotation $(\bar{T}, \bar{F})$ for those bands is farther from zero (Fig. 3). A secondary cause may be the small increase in standard deviation at higher concentrations. Although such a trend was not statistically significant for these data, it could be a major causative factor where variances increase greatly with concentration increases. In any case, none of the CRL estimates for TRL = 41.1 μg/L are acceptable because they all fall below the concentration of the 0.5 TRL standard, 20.5 μg/L (USATHAMA 1987). When TRL was 20.5 μg/L, four of the 15 estimates were slightly below the lowest standard, 10.3 μg/L. For TRL = 10.3 μg/L all estimates were acceptable.

2. The first truncation produced a reduction in CRL only for the highest TRL, but the second truncation reduced CRL estimates for all three TRLs. This effect was not very pronounced despite reduced extrapolation associated with lowering the value of $(\bar{T}, \bar{F})$, because calibration error was small compared to sample analysis error. Consequently, prediction bands did not exhibit much curvature.

3. The ranges of CRL estimates increased as TRL increased, presumably due to the greater effect of slight slope changes when long extrapolation was required. However, truncation, which reduces the concentration range, also produced an increase in the variability of CRL estimates. Here, the dominant influence is that the values for the second truncation are derived entirely from randomly varying data whereas the full data sets use the same response values for the highest standards in each iteration. Consequently, the variability in CRL estimates is more fairly represented by the values for the second truncation than for the full data sets.

The results are now at hand for a meaningful comparison of MDL and CRL estimates for manual peak height measurements of DNB concentrations. In Table 7 an MDL value of 6.84 μg/L was reported using the current EPA guidelines of $\alpha = 1\%$ and $\beta = 50\%$. This estimate was based on a pooled standard deviation but it assumed only 9 d.f. since it was to represent a typical value for one set of 10 replicate measurements. If only 6 d.f. were used in accordance with the minimum required by the protocol, very slight increases in MDL could result because of larger values of t. The current USA THAMA protocol based on duplicate measurements at each of five concentrations on four separate days and $\alpha = \beta = 5\%$ gave an average CRL of 8.4 μg/L when TRL was 10.3 μg/L (Table 8). If the MDL estimate was also based on $\alpha = \beta = 5\%$, then the value was 8.59 μg/L (Table 7). Clearly, these MDL and CRL estimates compare very favorably, especially considering that the CRL estimates include day-to-day variability and lack-of-fit of regression models, whereas the MDL estimates exclude these sources of uncertainty. However, CRL estimates increased significantly as TRL increased. In contrast, MDL estimates were quite insensitive to the concentration used for data generation in this particular analysis. It is concluded that the selection of an acceptably small TRL will yield CRL estimates that are very similar to MDL estimates.

Table 9. Mean found concentrations and standard deviations for GFAA estimates of Cu (edited data).

Data set	*Concentration (µg/L)*		*Degrees of freedom (d.f.)*	*Standard deviation (s)*
	Taken	*Found ($\bar{X}$)*		
Day 1	0.150	0.193	6	0.017
Day 2	0.150	0.143	6	0.027
Day 3	0.150	0.138	6	0.029
Day 4	0.150	0.147	6	0.046
Total*	0.150	0.155	27	0.037
Pooled†	—	—	24	0.031
Day 1	0.300	0.325	6	0.028
Day 2	0.300	0.283	6	0.031
Day 3	0.300	0.263	6	0.021
Day 4	0.300	0.281	6	0.029
Total*	0.300	0.288	27	0.035
Pooled†	—	—	24	0.028
Day 1	.600	0.621	6	0.026
Day 2	0.600	0.550	6	0.042
Day 3	0.600	0.551	6	0.042
Day 4	0.600	0.557	6	0.055
Total*	0.600	0.570	27	0.050
Pooled†	—	—	24	0.043
Day 1	.20	1.19	6	0.043
Day 2	1.20	1.08	5	0.050
Day 3	1.20	1.07	6	0.101
Day 4	1.20	1.05	6	0.070
Total*	1.20	1.10	27	0.087
Pooled†	—	—	24	0.070
Day 1	2.40	2.29	6	0.102
Day 2	2.40	2.20	6	0.089
Day 3	2.40	2.19	6	0.145
Day 4	2.40	2.36	6	0.168
Total*	2.40	2.26	27	0.141
Pooled†	—	—	24	0.130
Total**	3.00	2.85	7	0.081
Total**	6.00	6.01	7	0.268
Total**	12.00	12.06	7	0.817

* The total standard deviation is based on combining the data from all four days into a single set.
† The pooled standard deviation combines the four variances from each day but with day-to-day variations excluded.
** For the three highest concentrations where duplicates were run each day, only the total standard deviation is reported

It is also worth noting that there is no a priori requirement that $\alpha = \beta$ when generating CRL estimates. The required equations for computations with $\alpha \neq \beta$ are given in Wernimont (1985, p. 76). Similarly, the estimated COD and the CRL could be specified as suggested in the *Theory* section.

MDL estimates for Cu data set

Blank-corrected copper results are presented in Appendix C. One value was excluded as an outlier based on Dixon's test at a 5% significance level. Cochran's range comparison (Youden and Steiner 1975) for variance homogeneity of the five lowest concentrations demonstrated heterogeneity at the 5% significance level. When only the three lowest concentrations were similarly tested, variances were homogeneous.

Mean found concentrations and associated standard deviations for individual measurements are summarized in Table 9 in an arrangement analogous to the DNB results in Table 4. In general mean found concentrations were in reasonable agreement with taken concentrations although recoveries tended to be slightly low in some cases. Daily estimates of s for the three lowest concentrations ranged from 0.017 to 0.055 µg/L. Above a taken concentration of 0.600 µg/L, s increased substantially. In contrast to the DNB data, the pooled estimates of s were consistently smaller than the estimates based on the total data set, suggesting that day-to-day calibration variations were important. This tentative conclusion was confirmed via analysis of variance (ANOVA) which showed differences in daily mean responses ($\alpha = 0.05$) at several concentrations.

As with DNB results, MDLs were calculated according to EPA recommendations (Federal Register 1984) using the s values from sets of daily replicates. Since variance homogeneity was demonstrated for copper concentrations of 0.150, 0.300 and 0.600 µg/L, MDL estimates were obtained for each of these concentrations (Table 10). Once again, the highest concentration (0.600 µg/L) is slightly above the specified one to five times the estimated MDL but we chose to retain the values. The MDLs ranged from 0.052 to 0.172 µg/L, or omitting the 0.600-µg/L results, 0.052 to 0.145 µg/L. For the pooled standard deviations, the estimates only ranged from 0.070 to 0.107 µg/L. The mean MDL estimate for the day 1 data was 0.074 µg/L and the 95% confidence limits around this value, assuming 6 degrees of freedom, were 0.048 to 0.163 µg/L. Only one of the 12 MDL estimates for individual data sets is outside of this range. Of course, changes in risk assumptions will produce a pattern of variation in MDL estimates similar to that shown for DNB in Table 7.

We must also note that the copper MDLs should actually be increased by to account for the "furnace blank" corrections that were applied to all of these responses. However, since the current EPA procedure does not require this adjustment, it has been omitted here. Fortunately this omission does not invalidate MDL and CRL comparisons because

Table 10. Method detection limit (MDL)* estimates for graphite furnace atomic absorption measurements of copper concentrations. Estimates are based on *t*-values with $\alpha = 1\%$ and $\beta = 50\%$.

Conc taken	*MDL estimates (µg/L)*				
(µg/L)	*Day 1*	*Day 2*	*Day 3*	*Day 4*	*Pooled*
5.1	8.5	5.1	9.7	12.1	8.0
0.150	0.052	0.085	0.090	0.145	0.077
0.300	0.088	0.099	0.065	0.093	0.070
0.600	0.081	0.133	0.132	0.172	0.107

* MDL = $t_{0.99}(s)$.
$t_{0.99}$(d.f. = 6) = 3.143, $t_{0.99}$(d.f. = 24) = 2.492.

the latter estimates were also obtained using blank corrected data.

CRL estimates for Cu data set

CRL estimates were obtained as described in the discussion of DNB results except that all Cu responses were corrected for the "furnace blank" (Table 11). Trends in these values are as follows:

1. As TRL increased, CRL estimates also increased for both full and truncated data sets. Specifically, CRL ranged from 0.113 to 0.237 µg/L for TRL = 0.300 µg/L, from 0.238 to 0.523 µg/L for TRL = 0.600 µg/L, and from 0.365 to 1.32 for TRL = 1.20 µg/L. Corresponding means were 0.182, 0.350, and 0.768 µg/L (Table 11). The proportionate increase in Cu CRL is much greater than that observed for DNB with the same relative increase in TRL. With Cu, the increase is not primarily due to greater extrapolation from the centroid as TRL increases, but because the standard deviation is much larger for high than for low Cu concentration (Table 9). As noted in the discussion of DNB results, heterogeneous variances can cause this to be a dominant effect. Hence, the influence of the choice of TRL is much greater for Cu than for DNB.

2. The impact of variance heterogeneity is also strongly evident when truncation is employed. In contrast to DNB results where truncation produced only minor reductions in CRL estimates, the CRLs for Cu are dramatically reduced with successive truncation, especially for large TRL. Clearly the reduction of *s* associated with deletion of high concentration data accounts for most of this decrease in CRL estimates. It is also worth noting that the choice of TRL is extremely important because of the USAHTAMA requirement that 0.5 TRL becomes the CRL when the calculated CRL is below 0.5 TRL. For Cu, that restriction has only a minor effect for full data sets and after one truncation (although all five CRLs for one truncation and TRL = 0.600 µg/L are slightly below 0.300 µg/L). However, after the second truncation, 12 of 15 CRLs are below 0.5 TRL. Since truncation is routinely used, the CRL is effectively determined by the TRL unless an iterative procedure involving lower TRLs is employed or a very low TRL is chosen initially. In practice, every effort is made to choose a TRL as low as method capability permits, thereby minimizing this problem. Alternatively, the USA THAMA (1987) protocol could be altered to include standard concentrations of 0.25 TRL or 0.1 TRL. The same problem was evident with DNB but it was less serious than for Cu.

3. As noted for DNB, only the twice truncated data represent random results since the high concentration responses were the same in each iteration. Consequently, the variability of CRL estimates is best reflected by the values for the second truncation (Table 11). There was no apparent systematic dependence of this variability on the size of TRL.

Table 11. Certified reporting limit (CRL)* estimates for graphite furnace atomic absorption copper determinations.

Target reporting limit (µg/L)	*Full data set*	*Highest conc. deleted*	*Two highest conc. deleted*
0.300	0.237	0.230	0.187
0.300	0.222	0.195	0.113
0.300	0.190	0.155	0.145
0.300	0.200	0.179	0.139
0.300	0.197	0.173	0.171
	$\bar{X}$ = 0.209	$\bar{X}$ = 0.186	$\bar{X}$ = 0.151
	R† = 0.047	*R* = 0.075	*R* = 0.074
0.600	0.496	0.274	0.268
0.600	0.485	0.246	0.238
0.600	0.497	0.279	0.264
0.600	0.485	0.244	0.238
0.600	0.523	0.339	0.377
	$\bar{X}$ = 0.497	$\bar{X}$ = 0.276	$\bar{X}$ = 0.277
	R = 0.038	*R* = 0.095	*R* = 0.139
1.20	1.30	0.571	0.421
1.20	1.32	0.597	0.365
1.20	1.31	0.604	0.444
1.20	1.30	0.557	0.398
1.20	1.30	0.583	0.448
	$\bar{X}$ = 1.31	$\bar{X}$ = 0.582	$\bar{X}$ = 0.415
	R = 0.03	*R* = 0.047	*R* = 0.084

* Based on $\alpha = \beta = 5\%$ using USATHAMA (1987) computational protocol (Fig. 2).
† Range between high and low CRL estimate.

Let us now compare MDL with CRL for the Cu determinations. Based on a pooled standard deviation of 0.034 μg/L from the three lowest Cu concentrations and assuming 6 degrees of freedom in the estimate, the MDL = 0.107 μg/L according to current EPA guidelines (α = 1%, β = 50%). For α = β = 5% the corresponding MDL = 0.132 μg/L. The CRL estimate, also based on α = β = 5%, was 0.151 μg/L when TRL = 0.300 μg/L and the two highest concentrations were omitted by truncation. These two estimates compare quite favorably. With no truncation the CRL was 0.209 μg/L for TRL = 0.300 μg/L. As TRL increased, the CRL estimates rapidly escalated to produce values that were much larger (3–13 times) than MDL. In contrast, the CRL estimates for DNB only exceeded the MDL estimates by a factor of a little more than 2 in the worst case. This difference in behavior is a clear reflection of the effect of variance heterogeneity. When variances increase even in the low concentration region, as they do for Cu, both MDL and CRL estimates require very careful choice of concentrations used for data generation. However, the impact of a poor choice (too high a concentration) is more severe for CRL than for MDL estimates.

CONCLUSIONS AND RECOMMENDATIONS

1. A strong case has been made for using outlier tests and variance homogeneity tests to edit data. This process should only exclude a very small percentage of extreme responses, preferably less than 10%. Without editing, a few highly aberrant values can unreasonably distort an otherwise valuable data set.

2. For chromatography (and probably for many other techniques) integrator responses should be carefully examined by visual inspection. Otherwise occasional spurious responses will be obtained, primarily at very low analyte concentrations. Often this inspection requires that an analog chromatogram also be obtained in addition to the integrator output.

3. The magnitude of variability of repeat Method Detection Limit (MDL) estimates and Certified Reporting Limit (CRL) estimates reinforces our view that such descriptors are not fundamental parameters.

4. The influence of risk assumptions on MDL and CRL estimates has been demonstrated. It is recommended that proper attention must be given to type II errors (β risk or false negatives) because current procedures set this risk at 50%. To overcome this problem, it is suggested that the criterion or detection be used in conjunction with MDL or CRL (see *Theory* section). The choice of both α and β risks should be properly married to the problem at hand.

5. As Target Reporting Limits (TRL) increase, CRL estimates also increase. This behavior is most evident when variances increase with concentration, even in the low concentration range, as exemplified by graphite furnace atomic absorption determinations of copper. The possible advantages of weighted least squares should be examined here. This problem can also be minimized by choosing TRL as low as practical considerations allow.

6. Truncation of "found vs taken" curves lowers CRL estimates although the effect is much more profound in the presence of heterogeneous variances. Here too, weighted least squares might reduce this effect but adequate software would have to be found or produced. Even with homogeneous variances, truncation will still produce some decrease in CRL due to the reduction of extrapolation of confidence bands when the centroid values are made smaller.

7. For both DNB and Cu, the process of truncation often led to CRL estimates that were smaller than 0.5 TRL. Current USATHAMA requirements are that the CRL cannot be less than 0.5 TRL. In effect, the lower limit for CRL is predetermined at 0.5 TRL by the choice of TRL and the performance capability of the method may be discarded. To overcome this undesirable situation, it is recommended that either this requirement be eliminated the lowest solution concentration be either 0.1 TRL or 0.2 TRL, rather than 0.5 TRL. Ideally, this lower concentration would be added and the 0.5 TRL retained to accommodate the times when the TRL is close to the method capability and 0.1 TRL or 0.2 TRL will provide no response.

8. For systems with reasonably constant variance in the low concentration region (i.e. RP–HPLC determination of DNB), MDL and CRL estimates according to current risk assumptions show remarkably good agreement. However, when variance increases with concentration (Cu), CRL can be considerably larger than MDL. The magnitude of this discrepancy is directly related to the size of TRL. Consequently, to achieve reasonable coincidence of these estimates, TRL should always be assigned a value as low as is practical.

LITERATURE CITED

ACS Committee on Environmental Improvement (1980) Guidelines for data acquisition and data quality evaluation in environmental chemistry. *Analytical Chemistry*, **52**: 2242–2249.

Clayton, C.A., J.W. Hines and P.D. Elkins (1987) Detection limits with specified assurance probabilities. *Analytical Chemistry*, **59**: 2506–2514.

Currie, L.A. (1968) Limits for qualitative detection and quantitative determination. *Analytical Chemistry*, **40**: 586-593.

Currie, L.A., Ed. (1988) *Detection in Analytical Chemistry: Importance, Theory, and Practice.* Washington, D.C.: American Chemical Society Symposium Series 361.

Dixon, W.J. (1953) Processing data for outliers. *Biometrics*, March, 74–89.

Dixon, W.J. and F.J. Massey, Jr. (1969) *Introduction to Statistical Analysis.* New York: McGraw-Hill.

Federal Register (1984) Definition and procedure for the determination of the method detection limit. *Code of Federal Regulations*, Part 136, Appendix B, October 26.

Glaser, J.A., D.L. Forest, G.D. McKee, S.A. Quave and W.L. Budde (1981) Trace analyses for wastewaters. *Environmental Science and Technology*, **15**: 1426–1435.

Grubbs, F.E. (1969) Procedures for detecting outlying observations in samples. *Technometrics*, **11**(1): 1–21.

Hubaux, A. and G. Vos (1970) Decision and detection limits for linear calibration curves. *Analytical Chemistry*, **42**: 849–855.

Jenkins, T.F. and M.E. Walsh (1987) Development of an analytical method for explosive residues in soil. USA Cold Regions Research and Engineering Laboratory, CRREL Report 87-7.

Jenkins, T.F., P.H. Miyares and P.W. Schumacher (1988) Development of an improved method for the determination of nitroaromatics and nitramines in water. USA Cold Regions Research and Engineering Laboratory, Special Report 88-23.

Kirchmer, C.J. (1983) Quality control in water analyses. *Environmental Science and Technology*, **17** (4): 174A–181A.

Kirchmer, C.J. (1988) Estimation of detection limits for environmental analytical procedures. In *Detection in Analytical Chemistry: Importance, Theory and Practice.* (L.A. Currie, Ed.), Washington, D.C.: American Chemical Society Symposium Series 361, pp. 78-93.

Maskarinec, M.P. and S.K. Holladay (1987) Quality assurance/quality control in waste site characterization and remedial action. Oak Ridge National Laboratory, Oak Ridge, Tennessee, Final Report ORNL/TM-10600.

Oppenheimer, L., T.P. Capizzi, R.M. Weppelman and H. Mehta (1983) Determining the lowest limit of reliable assay measurement. *Analytical Chemistry*, **55**: 638–643.

USATHAMA (1987) USATHAMA QA program. USA Toxic and Hazardous Materials Agency, Aberdeen Proving Ground, Maryland, 2nd Ed.

Wernimont, G.T. (1985) *Use of Statistics to Develop and Evaluate Analytical Methods*, Association of Official Analytical Chemists, Arlington, Virginia.

Youden, W.J. and E.H. Steiner (1975) *Statistical Manual of the Association of Official Analytical Chemists: Statistical Techniques for Collaborative Tests.* Association of Official Analytical Chemists, Arlington, Virginia.

APPENDIX A: MANUAL PEAK HEIGHT MEASUREMENTS OF CHROMATOGRAMS USED TO ESTIMATE DNB CONCENTRATIONS

DNB conc taken (µg/L)	DNB conc (µg/L)			
	Day 1	Day 2	Day 3	Day 4
0.0	0.0	4.5	0.0	0.0
0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0
0.0	0.0	4.9	4.8	0.0
0.0	0.0	0.0	6.3	0.0
0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	6.9
0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0
5.1	7.2	5.4	13.6	0.0
5.1	6.3	4.0	7.5	7.0
5.1	6.3	5.6	2.7	9.5
5.1	3.6	5.8	6.1	0.0
5.1	7.2	9.4	5.2	11.3
5.1	10.5	4.5	9.7	8.9
5.1	8.1	7.1	7.0	5.1
5.1	2.7	8.5	2.5	6.2
5.1	0.0	8.5	10.0	0.0
5.1	4.5	6.0	9.3	9.1
10.3	10.8	10.5	14.0	9.7
10.3	11.4	11.6	11.8	11.5
10.3	13.0	12.9	9.3	10.2
10.3	10.8	10.5	11.8	9.1
10.3	11.9	11.6	12.7	15.3
10.3	11.7	13.4	9.3	13.5
10.3	9.2	11.6	11.3	10.6
10.3	7.8	11.4	7.5	8.0
10.3	10.1	16.1	13.6	11.5
10.3	9.4	13.6	14.0	11.5
20.5	22.9	20.7	23.1	14.0
20.5	22.4	21.4	21.7	20.1
20.5	20.8	23.0	22.2	23.9

DNB conc taken (µg/L)	DNB conc (µg/L)			
	Day 1	Day 2	Day 3	Day 4
20.5	19.7	24.7	20.8	23.0
20.5	20.6	23.0	22.9	19.4
20.5	22.0	20.5	22.9	16.9
20.5	20.4	21.0	23.1	25.2
20.5	22.0	20.3	**15.9***	17.7
20.5	23.5	18.3	21.1	24.1
20.5	24.0	25.0	21.1	23.3
41.1	44.4	45.0	43.9	45.5
41.1	41.0	42.1	43.0	37.9
41.1	39.5	46.6	46.4	39.7
41.1	41.2	43.9	40.8	36.5
41.1	40.8	45.3	**31.3***	42.4
41.1	43.0	46.2	38.7	**54.6***
41.1	43.0	37.7	43.4	41.4
41.1	39.9	39.2	39.2	43.4
41.1	39.5	40.4	40.5	38.8
41.1	40.8	44.8	38.7	40.3
82.2	77.3	84.5	83.8	**94.4***
82.2	84.3	80.7	78.8	81.6
82.2	**96.4***	79.6	83.1	78.9
82.2	81.1	81.2	82.0	80.7
82.2	82.0	76.7	84.0	84.2
82.2	81.8	81.2	**69.7***	81.3
82.2	80.9	76.3	80.2	79.3
82.2	80.5			
82.2	82.0			
82.2	81.4			
103.0	101.3	101.9	93.3	99.4
103.0	100.0	100.8	108.7	103.2
205.0	202.4	197.1	195.7	200.6
205.0	200.2	206.7	201.3	204.5
410.0	401.7	404.2	399.0	389.8
410.0	401.0	405.4	394.7	386.5

* Boldfaced values were outliers at the 5% significance level according to Dixon's Test.

APPENDIX B: INTEGRATOR PEAK HEIGHT MEASUREMENTS OF CHROMATOGRAMS USED TO ESTIMATE DNB CONCENTRATIONS

DNB conc taken (µg/L)	DNB conc (µg/L) Day 1	Day 2	Day 3	Day 4
0.0	0.0	11.9	6.4	18.4
0.0	0.0	0.0	11.3	0.0
0.0	0.0	0.0	6.3	0.0
0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	7.2
0.0	0.0	15.4	8.6	0.0
0.0	4.4	5.2	0.0	0.0
0.0	0.0	0.0	0.0	0.0
0.0	5.1	0.0	0.0	10.6
5.1	9.8	12.7	9.5	10.7
5.1	6.8	15.3	10.7	**40.4***
5.1	8.3	9.9	0.0	10.6
5.1	0.0	0.0	6.1	9.7
5.1	10.3	10.8	9.0	5.9
5.1	0.0	0.0	0.0	15.8
5.1	0.0	14.0	10.5	0.0
5.1	0.0	13.4	8.2	0.0
5.1	13.1	23.3	6.9	11.7
5.1	0.0	12.6	13.9	12.2
10.3	12.7	13.0	22.0	19.6
10.3	0.0	13.4	20.1	13.2
10.3	15.2	12.2	11.8	0.0
10.3	21.0	15.1	29.3	13.0
10.3	15.0	24.9	13.9	16.8
10.3	16.4	20.8	18.7	12.5
10.3	12.9	11.3	21.1	0.0
10.3	0.0	11.5	15.0	21.4
10.3	17.3	22.2	11.9	12.8
10.3	14.9	12.4	17.5	8.7
20.5	25.7	27.4	28.5	39.0
20.5	30.6	21.5	29.8	22.7
20.5	32.9	23.7	26.3	21.0
20.5	25.9	27.7	28.8	20.3

DNB conc taken (µg/L)	DNB conc (µg/L) Day 1	Day 2	Day 3	Day 4
20.5	21.8	**40.5†**	18.5	21.8
20.5	23.0	24.5	22.8	27.5
20.5	26.2	28.5	32.1	30.5
20.5	27.8	**6.1†**	20.3	35.5
20.5	**48.7***	20.0	21.1	29.1
20.5	24.4	27.1	19.4	26.7
41.1	41.2	58.7	55.2	35.6
41.1	43.9	47.2	47.3	42.0
41.1	41.8	47.7	53.9	47.5
41.1	42.8	48.4	41.1	48.3
41.1	47.6	44.0	31.7	58.2
41.1	**55.6***	56.3	38.5	50.0
41.1	44.6	48.5	59.2	38.6
41.1	42.2	47.7	52.6	65.3
41.1	43.2	48.8	44.2	47.0
41.1	42.8	45.0	51.7	52.5
82.2	83.8	97.6	82.2	84.7
82.2	89.7	100.0	79.2	89.4
82.2	101.6	82.4	86.7	91.4
82.2	93.5	95.7	79.4	90.8
82.2	87.2	81.0	86.8	85.1
82.2	84.7	89.5	74.7	79.5
82.2	91.1	76.5	81.8	89.8
82.2	80.3			
82.2	99.8			
82.2	86.5			
103.0	107.9	105.4	93.4	114.3
103.0	116.3	111.0	117.4	104.2
205.0	211.6	213.3	190.0	196.1
205.0	203.7	212.9	208.2	186.4
410.0	397.5	416.6	411.3	396.6
410.0	400.9	401.8	381.0	389.3

* Boldfaced values were outliers at the 5% significance level according to Dixon's Test.

† Boldfaced values were outliers at the 5% significance level according to Grubb's Test

APPENDIX C: GRAPHITE FURNACE ATOMIC ABSORPTION MEASUREMENTS OF COPPER CONCENTRATIONS

Conc taken (µg/L)	MDL Estimates (µg/L) Day 1	Day 2	Day 3	Day 4
0.150	0.193	0.133	0.146	0.207
0.150	0.193	0.162	0.146	0.177
0.150	0.193	0.133	0.146	0.177
0.150	0.222	0.133	0.116	0.147
0.150	0.193	0.133	0.176	0.086
0.150	0.193	0.090	0.086	0.086
0.150	0.164	0.105	0.146	0.147
0.300	0.337	0.331	0.267	0.268
0.300	0.308	0.275	0.297	0.298
0.300	0.366	0.303	0.267	0.328
0.300	0.337	0.275	0.267	0.237
0.300	0.280	0.303	0.267	0.268
0.300	0.308	0.247	0.237	0.269
0.300	0.337	0.247	0.237	0.298
0.600	0.654	0.614	0.568	0.660
0.600	0.625	0.586	0.568	0.539
0.600	0.625	0.501	0.598	0.539
0.600	0.596	0.501	0.568	0.600
0.600	0.596	0.558	0.478	0.539
0.600	0.654	0.558	0.508	0.509
0.600	0.596	0.530	0.568	0.509

Conc taken (µg/L)	MDL Estimates (µg/L) Day 1	Day 2	Day 3	Day 4
1.20	1.12	1.07	0.96	1.11
1.20	1.23	1.07	1.08	1.14
1.20	1.23	1.07	1.20	1.02
1.20	1.20	**0.78***	0.96	1.05
1.20	1.17	1.01	1.05	1.02
1.20	1.17	1.12	1.20	1.08
1.20	1.23	1.15	1.02	0.93
2.40	2.33	2.37	2.19	2.44
2.40	2.21	2.09	2.41	2.26
2.40	2.35	2.14	2.10	2.59
2.40	2.09	2.17	2.13	2.44
2.40	2.38	2.17	2.04	2.17
2.40	2.30	2.23	2.38	2.14
2.40	2.35	2.20	2.07	2.44
3.00	2.87	2.82	2.92	2.96
3.00	2.93	2.71	2.80	2.87
6.00	5.90	5.85	6.02	6.25
6.00	6.60	5.91	5.54	6.31
12.0	11.37	12.81	11.80	12.92
12.0	10.85	12.70	11.32	12.68

* Boldfaced value was an outlier at the 5% significance level according to Dixon's Test

U. S. GOVERNMENT PRINTING OFFICE 1989/600-057/02029